Title:

Genetic diversity on the human X chromosome does not support a strict pseudoautosomal boundary.

Authors and affiliations:

Daniel J. Cotter[1]*, Sarah M. Brotman[1]* and Melissa A. Wilson Sayres[1,2]

* These authors contributed equally.

[1] School of Life Sciences, Arizona State University, Tempe, Arizona, 85281 USA

[2] Center for Evolution and Medicine, The Biodesign Institute, Arizona State University, Tempe, Arizona, 85281 USA

Running title:

No strict pseudoautosomal boundary

Keywords: nucleotide diversity; pseudoautosomal region (PAR); X-transposed region (XTR);

sex chromosome evolution; recombination

Corresponding author:

Melissa A. Wilson Sayres

School of Life Sciences, Arizona State University

PO Box 874501

Tempe, AZ 85287-4501

p. 480-727-6366

e. melissa.wilsonsayres@asu.edu

Abstract

Unlike the autosomes, recombination between the X chromosome and Y chromosome often thought to be constrained to two small pseudoautosomal regions (PARs) at the tips of each sex chromosome. The PAR1 spans the first 2.7 Mb of the proximal arm of the human sex chromosomes, while the much smaller PAR2 encompasses the distal 320 kb of the long arm of each sex chromosome. In addition to the PAR1 and PAR2, there is a human-specific X-transposed region that was duplicated from the X to the Y. The X-transposed region is often not excluded from X-specific analyses, unlike the PARs, because it is not thought to routinely recombine. Genetic diversity is expected to be higher in recombining regions than in non-recombining regions because recombination reduces the effect of linked selection. In this study, we investigate patterns of genetic diversity in noncoding regions across the entire X chromosome of a global sample of 26 unrelated genetic females. We observe that genetic diversity in the PAR1 is significantly greater than the non-recombining regions (nonPARs). However, rather than an abrupt drop in diversity at the pseudoautosomal boundary, there is a gradual reduction in diversity from the recombining through the non-recombining region, suggesting that recombination between the human sex chromosomes spans across the currently defined pseudoautosomal boundary. In contrast, diversity in the PAR2 is not significantly elevated compared to the nonPAR, suggesting that recombination is not obligatory in the PAR2. Finally, diversity in the X-transposed region is higher than the surrounding nonPAR regions, providing evidence that recombination may occur with some frequency between the X and Y in the XTR.

Introduction

The human sex chromosomes, X and Y, were previously an indistinguishable pair of autosomes, but within the last 180-210 million years, the ancestral pair diverged into two distinct chromosomes of tremendously different gene content and function [1,2]. The human sex chromosomes are composed of an older X-conserved region, shared across all therian (marsupial and eutherian) mammals [3,4], and a younger X- and Y-added region: autosomal sequence that was translocated to the X and Y chromosomes in the common ancestor of eutherian mammals approximately 80-130 million years ago [5]. The differentiation of the X and Y is hypothesized to have occurred after a series of Y-specific inversions that suppressed X-Y recombination [6-10]. In the absence of homologous recombination, the Y chromosome has lost nearly 90% of the genes that were on the ancestral sex chromosomes [11-13]. Today the human X chromosome and Y chromosome share two pseudoautosomal regions (PAR1 and PAR2) at the ends of the chromosomes that continue to undergo homologous X-Y recombination [6]. The PAR1 spans the first 2.7 Mb of the proximal arm of the human sex chromosomes [11] and contains genes from the ancient X- and Y-added region translocation. The PAR1 is separated from the non-recombining (nonPAR) region on the Y chromosome by a Y-specific inversion that is hypothesized to suppress X-Y recombination at this pseudoautosomal boundary [10]. A functional copy of the XG gene spans the human pseudoautosomal boundary on the X chromosome [14], but is interrupted on the Y chromosome by a Y-specific inversion [15]. In contrast to this mechanism for PAR formation, the 320kb human-specific PAR2 resulted from at least two duplications from the X chromosome to the terminal end of the Y chromosome [16].

Genes located in the PAR1 have important functions in all humans. Although genes on one X chromosome in 46,XX individuals are silenced via a process called X-inactivation [17], which evolved in response to loss of homologous gene content on the Y chromosome [13], all 24 genes in the PAR1 escape inactivation [11,18,19] (Supplementary Table 1). For example, one gene in the PAR1, SHOX1, plays an important role in long bone growth and skeletal formation [20–22]. Consequence of SHOX1 disruption include short stature, skeletal deformities, Leri-Weill syndrome, and phenotypes associated with Turner syndrome (45,X) [20]. ASMT, another gene located in the PAR1, is involved in the synthesis of melatonin, and is thought to be connected with psychiatric disorders, including bipolar affective disorder [23].

The suggested function of the pseudoautosomal regions is to assist in chromosome pairing and segregation [24]. It has been proposed, in humans and great apes, that crossover events are mandatory during male meiosis [25–27]. The analyses of human sperm suggest that a deficiency in recombination in PAR1 is significantly correlated with the occurrence of nondisjunction and results in Klinefelter syndrome (47, XXY) [28]. Deletions in the PAR1 are shown to lead to short stature, which is correlated with Turner Syndrome [29]. Further, the male sex-determining gene on the Y chromosome (SRY) is proximal to the PAR1 on the short arm of the Y. SRY can be translocated from the Y to the X during incongruent crossover events between the paternal PAR1s, resulting in SRY-positive XX males [30], or, more rarely, true hermaphroditism [31]. The chances that XX individuals will inherit a copy of the SRY gene during male meiosis are restricted by reduced recombination at the PAR1 boundary [32].

5

Previous studies estimate that the recombination rate is approximately twenty times the genome average in the PAR1 [26] and about five times the genome average in the PAR2 [33], likely because recombination events in XY individuals are restricted to the pseudoautosomal sequences, with the exception of possible gene conversion in regions outside of the PARs [34]. In addition to the PAR1 and PAR2, where recombination is known to occur between the X and Y, there is an X-transposed region (XTR) that was duplicated from the X to the Y in humans after human-chimpanzee divergence [11,12]. Already, the XTR has incurred several deletions, and an inversion, but maintains 98.78% homology between the X and Y [11], and retains two genes with functional X-linked and Y-linked homologs [12].

Genetic diversity is expected to be higher in the pseudoautosomal regions than the remainder of the sex chromosomes for several reasons. First, recombination can unlink alleles affected by selection from nearby sites, reducing the effects of background selection and genetic hitchhiking on reducing genetic diversity [35,36,3662]. Second, the effective size of the pseudoautosomal regions of the sex chromosomes should be larger (existing in two copies in all individuals) than the non-recombining region of the X chromosome, which exists in two copies in genetic females and only one copy in genetic males. Finally, genetic diversity may be higher in pseudoautosomal regions than in regions that do not recombine in both sexes if recombination increases the local mutation rate [37–39].

Studies of human population genetic variation often compare diversity on the X chromosome with diversity on the autosomes to make inferences about sex-biased human demographic history [404142]. Typically the PAR1 and PAR2 are filtered out of these studies, at the defined

pseudoautosomal boundaries, and the XTR is not filtered out. However, patterns of diversity across the entire human X chromosome, including transitions across the PARs and XTR have not been investigated to justify these common practices. In this study we investigate patterns of genetic diversity and divergence across the entire human X chromosome including the pseudoautosomal regions of twenty-six unrelated (46, XX) individuals sequenced to an average depth of 80X coverage by CompleteGenomics [44].

Results and Discussion

Human X-linked nucleotide diversity is high in the PAR1 but not PAR2.

We observe that uncorrected diversity is three times higher in the PAR1 than in the nonPAR regions, while uncorrected diversity in the PAR2 is not significantly greater than that in the nonPAR (Table 1; Figure 1; Supplementary Figure 1). We studied noncoding regions across the entire X chromosome, filtering out annotated genes, to minimize the effect of selection, but due to their small sizes, could not filter out regions far from genes in the PARs or XTR (Methods). However, mutation rate variation across the X chromosome may account for variable levels of diversity observed in the PAR and nonPAR regions. To correct for different mutation rates across the X chromosome, we normalize each region by divergence computed in that same region, with regions of high divergence filtered out, computed pairwise between human-chimpanzee, human-macaque, human-dog, or human-mouse (Supplementary Figure 2). Similar to the uncorrected diversity values, when we correct for mutation rate using any of these divergence values, we observe higher nucleotide diversity across humans in the PAR1 and PAR2 regions relative to the nonPAR regions, with diversity being significantly higher in the PAR1 than nonPAR (with XTR removed), and not significantly different between the PAR2 and nonPAR (Figure 1; Table 1; Supplementary Figure 1).

7

Curiously, human-chimpanzee and human-macaque divergence are quite high in the PAR1 relative to the nonPAR in a pattern that does not reflect diversity (Figure 1; Table 1). This results, predominantly, due to high inter-species divergence in the PAR1, and near the pseudoautosomal boundary (Supplementary Figure 3; Supplementary Figure 4). However, human-dog divergence roughly parallels uncorrected human diversity (Figure 1). Alignments between the human and mouse in PAR1 are unavailable.

Further, significantly elevated diversity in the PAR1 relative to the nonPAR cannot be attributed solely to mutation rate variation across the X chromosome, as the pattern remains after correction for divergence in each region (Figure 1; Table 1; Supplementary Figure 1). The pattern we observe is consistent with several processes, including: selection reducing variation more at linked sites in the nonPAR region than in the PAR1 due to reduced rates of recombination in the nonPAR relative to the PAR, or as a result of stronger drift in the nonPAR due to a smaller effective population size.

That we do not observe significantly elevated diversity in the PAR2 relative to the nonPAR is consistent with reports that the PAR2 undergoes X-Y recombination less frequently than the PAR1 [48], and supports assertions that in humans only one chiasma per chromosome is needed for proper segregation, rather than one per chromosome arm [49].

8

Diversity is significantly higher in the X-transposed region than the non-pseudoautosomal region.

Curiously, in addition to elevated rates of diversity in the previously described PAR1 and PAR2, we also observe that diversity is significantly higher in the recent X-transposed region, XTR, than in the nonPAR (Table 1; Supplementary Figure 1). This increased diversity cannot be attributed to mis-mapping between the X and Y as we only analyzed individuals with two X chromosomes (Methods). High diversity in the XTR contrasts with initial suggestions that there is no X-Y recombination in the XTR [12], and is consistent with recent reports of X-Y recombination in some human populations in this region [46].

Given the large size of the nonPAR region, and the small size of the XTR, 5Mb[11], one may wonder whether removing the XTR would make a difference to measured levels of diversity across the human X chromosome. The raw diversity of the nonPAR including the XTR, measured as π, is 0.000512 while the raw diversity of the nonPAR excluding the XTR is 0.000506 (Table 1). Removal of the XTR does decrease both estimates of diversity and divergence in the nonPAR, and its inclusion in studies of X-specific diversity will affect inferences made when comparing X-linked and autosomal variation [50–53].

Pseudoautosomal boundaries cannot be inferred from patterns of diversity.

 Recombination between the X and Y is expected to be suppressed at the pseudoautosomal boundary, where X-Y sequence homology diverges due to a Y-specific inversion [10,14,15]. If diversity correlates highly with recombination rate, and X-Y recombination is strictly suppressed in the nonPAR after the pseudoautosomal boundary, then diversity is expected to drop sharply

9

between the PAR1 and nonPAR. However, when we analyze patterns of human diversity in windows across the X chromosome (Methods), we do not observe an abrupt shift in the level of diversity between the PAR1 and nonPAR regions (Figure 1). The lack of an observable pseudoautosomal boundary based on diversity is clear whether small or large (100kb or 1Mb), or overlapping or non-overalapping windows are used (Methods; Supplementary Figure 7). The history of gene conversion between the sex chromosomes may contribute to the increased diversity levels[54] in the nonPAR side of the Y-specific inversion that marks the pseudoautosomal boundary. Human diversity uncorrected for divergence decreases from the proximal end of the PAR1 through the pseudoautosomal boundary and well into the nonPAR. A sex-specific map of the PAR1 found that male recombination is higher near the telomeres, and decreases near the pseudoautosomal boundary, while, in contrast, the female recombination rate reported in the same study in the PAR1 is fairly flat throughout the region and increases near the pseudoautosomal boundary [55]. Thus, genetic diversity uncorrected for divergence in the PAR1 appears to correlate with the male recombination rate. Curiously, however, a previous study of recombination rate in the PAR1 reported an increase in the female (but not the male) recombination rate near the proximal end of the PAR1 [56]. Thus, potentially both male and female recombination rates contribute to the linear decrease in diversity observed in the PAR1 from the proximal end of the X through the pseudoautosomal boundary. Although not yet mapped, when the data becomes available, it will be useful to compare patterns of diversity with sex-specific recombination maps across the entire X chromosome.

Conclusions

10

We show that diversity is indeed higher in the pseudoautosomal regions, and lower in the regions of the X that are not known to recombine in males (nonPAR). Diversity in the PAR1 is significantly higher than in the nonPAR regardless of normalizing the diversity with divergence between human and either chimpanzee, macaque, or dog to correct for mutation rate (Table 1; Figure 1). Diversity was also normalized with divergence from the mouse, but there is not alignment between human and mouse in the PAR1, due to a different evolutionary origin in the PAR1 and no common pseudoautosomal genes are shared between them [45]. We observe that diversity is lower in the PAR2 than expected, and is not significantly different from the nonPAR region. We also show that diversity is elevated in the XTR above other nonPAR regions, verifying recent observations that the region may still undergo homologous recombination between the X and Y chromosomes [46]. Finally, when analyzing patterns of genetic diversity in windows across the human X chromosome we find that there is no strict boundary, based solely on the levels of diversity, between the recombining and putatively non-recombining regions, which could be attributed to the evolutionary shift in the PAB over time, extending the PAR1 due to a PAR1 length polymorphism [47]. This could also suggest that non-homologous recombination at the pseudoautosomal boundaries may be common.

Our observations of patterns of diversity across regions of the human X chromosome with variable levels of recombination are consistent with previous reports that diversity and divergence are correlated with recombination rate in humans across the genome [38] , and explicitly in the PAR1 [57]. Elevated levels of diversity in the XTR suggest that, consistent with a recent report [46], this region may frequently undergo X-Y recombination. Curiously, we did not find a significant elevation in diversity in the PAR2, that, in agreement with its unusual evolution

11

[16], indicate that it rarely recombines between X and Y during meiosis. Further, the lack of a clear differentiation in diversity between the PAR1 and nonPAR suggests that recombination suppression between X and Y is still an actively evolving process in humans, as in other species [58]. This is consistent with evidence that the position of the pseudoautosomal boundary varies across mammals [59–62]. There is even evidence of polymorphism in the pseudoautosomal boundary in a pedigree analysis of a paternally inherited X chromosome in humans [47]. Taken together with previous inferences about the variation in pseudoautosomal boundaries, our observations suggest that assumptions should not be made of a strict suppression of X-Y recombination at the proposed human pseudoautosomal boundary.

Methods

We analyzed X chromosomes from twenty-six unrelated (46, XX) individuals sequenced by CompleteGenomics [44] (Supplementary Table 4). Sites were filtered requiring data be present (monomorphic or variable) in all twenty-six samples. Human-chimpanzee (hg19-panTro4), human-macaque (hg19-rheMac3), human-dog (hg19-canFam3), and human-mouse (hg19-mm10) alignments were extracted from the UCSC genome browser [63]. We curated the human-chimpanzee and human-macaque alignments to filter out segments that included autosomal sequence aligning to the X chromosome (Supplementary Table 2; Supplementary Figure 5; Supplementary Figure 6); These alignments were visualized using the gmaj software [64]. Additionally, we observed several regions across the X chromosome exhibiting heightened divergence between the human and the chimpanzee or the human and the macaque (Supplementary Figure 3; Supplementary Figure 4). Upon further inspection, these regions often contain multi-copy gene families that could lead to mismapping (Supplementary Table 2).

Divergence estimates were similar with and without these regions (Supplementary Table 3), and here we present results with these regions of high divergence near multi-copy gene families excluded. Some regions of the X chromosome exhibit extremely high inter-species divergence (Supplementary Figure 3; Supplementary Figure 4; Supplementary Table 2). These regions were excluded from this analysis, but we reach the same conclusions described above whether they are included or excluded.

We used the Galaxy tools [65] to filter out regions that could cause potential sequence misalignments and regions defined by the UCSC genome browser [63] that may be subject to selection: RefSeq genes, simple repeats, and repetitive elements. We attempted to filter out noncoding regions near genes, but doing so would leave very little analyzable sequence in the PAR1 and PAR2 regions.

We measured the diversity between the sequences as π, the average pairwise nucleotide differences per site between all sequences in the sample.

$$\pi = \frac{2}{L} \frac{k}{(k-1)} \sum_{i=1}^{k} p_i \sum_{j>i}^{k} p_j d_{ij}$$

L represents the number of called sites, k represents the number of DNA sequences, p_i and p_j are the frequencies of the corresponding alleles i and j, and d_{ij} is the number of sites containing nucleotide differences. Diversity was calculated within each specific region (PAR1, PAR2, XTR, nonPAR with XTR, and nonPAR without XTR), as well as across sliding and non-overlapping windows. We generated window interval files across the human X chromosome with Galaxy tools [65], and conducted analysis, in four sets of windows (Supplementary Figure 7): 1) in a 1 Mb

13

non-overlapping window, 2) a 1 Mb window with 100 kb sliding start positions, 3) a 100 kb non-overlapping window, and 4) a 100 kb window with 10 kb sliding start positions.

We similarly calculated human-chimpanzee, human-macaque, human-dog, and human-mouse species divergence along the X chromosome, in each of the four regions and in the same windows previously described. To normalize the data, π values were divided by the observed divergence within the same interval.

Chromosome X was divided into windows that were permutated without replacement 10,000 times to assess significant differences between diversity in each region (PAR1, nonPAR, XTR, PAR2). This analysis was repeated for uncorrected diversity and diversity corrected for human-chimpanzee, human-macaque, human-dog, and human-mouse divergence values. Empirical *P*-values were calculated by comparing differences between permutated samples with differences in observed diversity between each pair of regions. The negative correlation along the pseudoautosomal boundary was tested using linear regressions across 100kb windows covering a total of 3Mb for each regression (30 windows), shifting the window by 100 kb systematically (Figure 2). Each regression was analyzed for significance ($p < 0.05$) with all data points occurring before the first non-significant window being included in the significant dataset. All of the graphs were produced using R version 3.1.2 [66]. All codes used for this project can be found at: https://github.com/WilsonSayresLab/PARdiversity.

References

1. Mikkelsen, T.S., Wakefield, M.J., Aken, B., Amemiya, C.T., Chang, J.L., Duke, S., Garber, M., Gentles, A.J., Goodstadt, L., Heger, A., et al. (2007). Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences. Nature *447*, 167–177.

2. Rens, W., O'Brien, P.C., Grützner, F., Clarke, O., Graphodatskaya, D., Tsend-Ayush, E., Trifonov, V.A., Skelton, H., Wallis, M.C., Johnston, S., et al. (2007). The multiple sex chromosomes of platypus and echidna are not completely identical and several share homology with the avian Z. Genome Biol. *8*, R243.

3. Watson, J.M., Spencer, J.A., Riggs, A.D., and Graves, J.A. (1990). The X chromosome of monotremes shares a highly conserved region with the eutherian and marsupial X chromosomes despite the absence of X chromosome inactivation. Proc. Natl. Acad. Sci. U. S. A. *87*, 7125–7129.

4. Glas, R., Graves, J.A.M., Toder, R., Ferguson-Smith, M., and O'Brien, P.C. (1999). Cross-species chromosome painting between human and marsupial directly demonstrates the ancient region of the mammalian X. Mamm. Genome *10*, 1115–1116.

5. Waters, P.D., Duffy, B., Frost, C.J., Delbridge, M.L., and Graves, J.A. (2001). The human Y chromosome derives largely from a single autosomal region added to the sex chromosomes 80-130 million years ago. Cytogenet. Cell Genet. *92*, 74–79.

6. Lahn, B.T., and Page, D.C. (1999). Four Evolutionary Strata on the Human X Chromosome. Science *286*, 964–967.

7. Marais, G., and Galtier, N. (2003). Sex chromosomes: how X-Y recombination stops. Curr. Biol. *13*, R641–R643.

8. Wilson, M.A., and Makova, K.D. (2009). Evolution and Survival on Eutherian Sex Chromosomes. PLoS Genet *5*, e1000568.

9. Lemaitre, C., Braga, M.D.V., Gautier, C., Sagot, M.-F., Tannier, E., and Marais, G.A.B. (2009). Footprints of Inversions at Present and Past Pseudoautosomal Boundaries in Human Sex Chromosomes. Genome Biol. Evol. *1*, 56–66.

10. Pandey, R.S., Sayres, M.A.W., and Azad, R.K. (2013). Detecting evolutionary strata on the human X chromosome in the absence of gametologous Y-linked sequences. Genome Biol. Evol. evt139.

11. Ross, M.T., Grafham, D.V., Coffey, A.J., Scherer, S., McLay, K., Muzny, D., Platzer, M., Howell, G.R., Burrows, C., Bird, C.P., et al. (2005). The DNA sequence of the human X chromosome. Nature *434*, 325–337.

12. Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., et al. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature *423*, 825–837.

13. Sayres, M.A.W., and Makova, K.D. (2013). Gene Survival and Death on the Human Y Chromosome. Mol. Biol. Evol. *30*, 781–787.

14. Yi, S., Summers, T.J., Pearson, N.M., and Li, W.-H. (2004). Recombination Has Little Effect on the Rate of Sequence Divergence in Pseudoautosomal Boundary 1 Among Humans and Great Apes. Genome Res. *14*, 37–43.

15. Ellis, N., Taylor, A., Bengtsson, B.O., Kidd, J., Rogers, J., and Fellow, P.G. (1990). Population structure of the human pseudoautosomal boundary. Nature *344*, 663–665.

16. Charchar, F.J., Svartman, M., El-Mogharbel, N., Ventura, M., Kirby, P., Matarazzo, M.R., Ciccodicola, A., Rocchi, M., D'Esposito, M., and Graves, J.A.M. (2003). Complex Events in the Evolution of the Human Pseudoautosomal Region 2 (PAR2). Genome Res. *13*, 281–286.

17. Carrel, L., and Willard, H.F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature *434*, 400–404.

18. Perry, J., Palmer, S., Gabriel, A., and Ashworth, A. (2001). A Short Pseudoautosomal Region in Laboratory Mice. Genome Res. *11*, 1826–1832.

19. Helena Mangs, A., and Morris, B.J. (2007). The Human Pseudoautosomal Region (PAR): Origin, Function and Future. Curr. Genomics *8*, 129–136.

20. Rao, E., Blaschke, R.J., Marchini, A., Niesler, B., Burnett, M., and Rappold, G.A. (2001). The Leri–Weill and Turner syndrome homeobox gene SHOX encodes a cell-type specific transcriptional activator. Hum. Mol. Genet. *10*, 3083–3091.

21. Tsuchiya, T., Shibata, M., Numabe, H., Jinno, T., Nakabayashi, K., Nishimura, G., Nagai, T., Ogata, T., and Fukami, M. (2014). Compound heterozygous deletions in pseudoautosomal region 1 in an infant with mild manifestations of langer mesomelic dysplasia. Am. J. Med. Genet. A. *164*, 505–510.

22. Benito-Sanz, S., Aza-Carmona, M., Rodríguez-Estevez, A., Rica-Etxebarria, I., Gracia, R., Campos-Barros, A., and Heath, K.E. (2012). Identification of the first PAR1 deletion encompassing upstream SHOX enhancers in a family with idiopathic short stature. Eur. J. Hum. Genet. EJHG *20*, 125–127.

23. Flaquer, A., Jamra, R.A., Etterer, K., Díaz, G.O., Rivas, F., Rietschel, M., Cichon, S., Nöthen, M.M., and Strauch, K. (2010). A new susceptibility locus for bipolar affective disorder in PAR1 on Xp22.3/Yp11.3. Am. J. Med. Genet. B Neuropsychiatr. Genet. *153B*, 1110–1114.

24. Kauppi, L., Barchi, M., Baudat, F., Romanienko, P.J., Keeney, S., and Jasin, M. (2011). Distinct properties of the XY pseudoautosomal region crucial for male meiosis. Science *331*, 916–920.

25. Rouyer, F., Simmler, M.-C., Johnsson, C., Vergnaud, G., Cooke, H.J., and Weissenbach, J. (1986). A gradient of sex linkage in the pseudoautosomal region of the human sex chromosomes. Nature *319*, 291–295.

26. Lien, S., Szyda, J., Schechinger, B., Rappold, G., and Arnheim, N. (2000). Evidence for Heterogeneity in Recombination in the Human Pseudoautosomal Region: High Resolution Analysis by Sperm Typing and Radiation-Hybrid Mapping. Am. J. Hum. Genet. *66*, 557–566.

27. Kauppi, L., Jasin, M., and Keeney, S. (2012). The tricky path to recombining X and Y chromosomes in meiosis. Ann. N. Y. Acad. Sci. *1267*, 18–23.

28. Shi, Q., Spriggs, E., Field, L.L., Rademaker, A., Ko, E., Barclay, L., and Martin, R.H. (2002). Absence of Age Effect on Meiotic Recombination between Human X and Y Chromosomes. Am. J. Hum. Genet. *71*, 254–261.

29. Rao, E., Weiss, B., Fukami, M., Rump, A., Niesler, B., Mertz, A., Muroya, K., Binder, G., Kirsch, S., Winkelmann, M., et al. (1997). Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. Nat. Genet. *16*, 54–63.

30. Page, D.C., de la Chapelle, A., and Weissenbach, J. (1985). Chromosome Y-specific DNA in related human XX males. Nature *315*, 224–226.

31. Abbas, N., McElreavey, K., Leconiat, M., Vilain, E., Jaubert, F., Berger, R., Nihoul-Fekete, C., Rappaport, R., and Fellous, M. (1993). Familial case of 46,XX male and 46,XX true hermaphrodite associated with a paternal-derived SRY-bearing X chromosome. Comptes Rendus Académie Sci. Sér. III Sci. Vie *316*, 375–383.

32. Fukagawa, T., Nakamura, Y., Okumura, K., Nogami, M., Ando, A., Inoko, H., Saitou, N., and Ikemura, T. (1996). Human Pseudoautosomal Boundary-like Sequences: Expression and Involvement in Evolutionary Formation of the Present-day Pseudoautosomal Boundary of Human Sex Chromosomes. Hum. Mol. Genet. *5*, 23–32.

33. Filatov, D.A., and Gerrard, D.T. (2003). High mutation rates in human and ape pseudoautosomal genes. Gene *317*, 67–77.

34. Rosser, Z.H., Balaresque, P., and Jobling, M.A. (2009). Gene Conversion between the X Chromosome and the Male-Specific Region of the Y Chromosome at a Translocation Hotspot. Am. J. Hum. Genet. *85*, 130–134.

35. Charlesworth, B. (2012). The Effects of Deleterious Mutations on Evolution at Linked Sites. Genetics *190*, 5–22.

36. Vicoso, B., and Charlesworth, B. (2006). Evolution on the X chromosome: unusual patterns and processes. Nat. Rev. Genet. *7*, 645–653.

37. Huang, S.-W., Friedman, R., Yu, N., Yu, A., and Li, W.-H. (2005). How Strong Is the Mutagenicity of Recombination in Mammals? Mol. Biol. Evol. *22*, 426–431.

38. Hellmann, I., Ebersberger, I., Ptak, S.E., Pääbo, S., and Przeworski, M. (2003). A Neutral Explanation for the Correlation of Diversity with Recombination Rates in Humans. Am. J. Hum. Genet. *72*, 1527–1535.

39. Perry, J., and Ashworth, A. (1999). Evolutionary rate of a gene affected by chromosomal position. Curr. Biol. *9*, 987–S3.

40. Arbiza, L., Gottipati, S., Siepel, A., and Keinan, A. (2014). Contrasting X-Linked and Autosomal Diversity across 14 Human Populations. Am. J. Hum. Genet. *94*, 827–844.

41. Hammer, M.F., Mendez, F.L., Cox, M.P., Woerner, A.E., and Wall, J.D. (2008). Sex-Biased Evolutionary Forces Shape Genomic Patterns of Human Diversity. PLoS Genet *4*, e1000202.

42. Gottipati, S., Arbiza, L., Siepel, A., Clark, A.G., and Keinan, A. (2011). Analyses of X-linked and autosomal genetic variation in population-scale whole genome sequencing. Nat. Genet. *43*, 741–743.

43. Wilson Sayres, M.A., Lohmueller, K.E., and Nielsen, R. (2014). Natural Selection Reduced Diversity on Human Y Chromosomes. PLoS Genet *10*, e1004064.

44. Drmanac, R., Sparks, A.B., Callow, M.J., Halpern, A.L., Burns, N.L., Kermani, B.G., Carnevali, P., Nazarenko, I., Nilsen, G.B., Yeung, G., et al. (2010). Human Genome Sequencing Using Unchained Base Reads on Self-Assembling DNA Nanoarrays. Science *327*, 78–81.

45. Gianfrancesco, F., Sanges, R., Esposito, T., Tempesta, S., Rao, E., Rappold, G., Archidiacono, N., Graves, J.A.M., Forabosco, A., and D'Urso, M. (2001). Differential Divergence of Three Human Pseudoautosomal Genes and Their Mouse Homologs: Implications for Sex Chromosome Evolution. Genome Res. *11*, 2095–2100.

46. Veerappa, A.M., Padakannaya, P., and Ramachandra, N.B. (2013). Copy number variation-based polymorphism in a new pseudoautosomal region 3 (PAR3) of a human X-chromosome-transposed region (XTR) in the Y chromosome. Funct. Integr. Genomics *13*, 285–293.

47. Mensah, M.A., Hestand, M.S., Larmuseau, M.H.D., Isrie, M., Vanderheyden, N., Declercq, M., Souche, E.L., Van Houdt, J., Stoeva, R., Van Esch, H., et al. (2014). Pseudoautosomal Region 1 Length Polymorphism in the Human Population. PLoS Genet. *10*,.

48. Flaquer, A., Rappold, G.A., Wienker, T.F., and Fischer, C. (2008). The human pseudoautosomal regions: a review for genetic epidemiologists. Eur. J. Hum. Genet. *16*, 771–779.

49. Fledel-Alon, A., Wilson, D.J., Broman, K., Wen, X., Ober, C., Coop, G., and Przeworski, M. (2009). Broad-Scale Recombination Patterns Underlying Proper Disjunction in Humans. PLoS Genet *5*, e1000658.

50. Arbiza, L., Gottipati, S., Siepel, A., and Keinan, A. (2014). Contrasting X-Linked and Autosomal Diversity across 14 Human Populations. Am. J. Hum. Genet. *94*, 827–844.

51. Keinan, A., and Reich, D. (2010). Can a Sex-Biased Human Demography Account for the Reduced Effective Population Size of Chromosome X in Non-Africans? Mol. Biol. Evol. *27*, 2312–2321.

52. Wilson Sayres, M.A., Lohmueller, K.E., and Nielsen, R. (2014). Natural Selection Reduced Diversity on Human Y Chromosomes. PLoS Genet *10*, e1004064.

53. Gottipati, S., Arbiza, L., Siepel, A., Clark, A.G., and Keinan, A. (2011). Contrasting human X-linked and autosomal variation in population-scale whole genome sequencing. Nat. Genet. *43*, 741–743.

54. Trombetta, B., Sellitto, D., Scozzari, R., and Cruciani, F. (2014). Inter- and Intraspecies Phylogenetic Analyses Reveal Extensive X–Y Gene Conversion in the Evolution of Gametologous Sequences of Human Sex Chromosomes. Mol. Biol. Evol. *31*, 2108–2123.

55. Hinch, A.G., Altemose, N., Noor, N., Donnelly, P., and Myers, S.R. (2014). Recombination in the Human Pseudoautosomal Region PAR1. PLoS Genet. *10*,.

56. Henke, A., Fischer, C., and Rappold, G.A. (1993). Genetic map of the human pseudoautosomal region reveals a high rate of recombination in female meiosis at the Xp telomere. Genomics *18*, 478–485.

57. Bussell, J.J., Pearson, N.M., Kanda, R., Filatov, D.A., and Lahn, B.T. (2006). Human polymorphism and human–chimpanzee divergence in pseudoautosomal region correlate with local recombination rate. Gene *368*, 94–100.

58. Bergero, R., and Charlesworth, D. (2009). The evolution of restricted recombination in sex chromosomes. Trends Ecol. Evol. *24*, 94–102.

59. Raudsepp, T., and Chowdhary, B.P. (2008). The horse pseudoautosomal region (PAR): characterization and comparison with the human, chimp and mouse PARs. Cytogenet. Genome Res. *121*, 102–109.

60. Raudsepp, T., Das, P.J., Avila, F., and Chowdhary, B.P. (2012). The Pseudoautosomal Region and Sex Chromosome Aneuploidies in Domestic Species. Sex. Dev. *6*, 72–83.

61. White, M.A., Ikeda, A., and Payseur, B.A. (2012). A pronounced evolutionary shift of the pseudoautosomal region boundary in house mice. Mamm. Genome Off. J. Int. Mamm. Genome Soc. *23*, 454–466.

62. Otto, S.P., Pannell, J.R., Peichel, C.L., Ashman, T.-L., Charlesworth, D., Chippindale, A.K., Delph, L.F., Guerrero, R.F., Scarpino, S.V., and McAllister, B.F. (2011). About PAR: The distinct evolutionary dynamics of the pseudoautosomal region. Trends Genet. *27*, 358–367.

63. Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., et al. (2015). The UCSC Genome Browser database: 2015 update. Nucleic Acids Res. *43*, D670–D681.

64. Blanchette, M., Kent, W.J., Riemer, C., Elnitski, L., Smit, A.F.A., Roskin, K.M., Baertsch, R., Rosenbloom, K., Clawson, H., Green, E.D., et al. (2004). Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res. *14*, 708–715.

65. Blankenberg, D., Taylor, J., Nekrutenko, A., and Team, T.G. (2011). Making whole genome multiple alignments usable for biologists. Bioinformatics *27*, 2426–2428.

66. R Core Team (2015). R: A Language and Environment for Statistical Computing (Vienna, Austria: R Foundation for Statistical Computing).

Tables

Region	Uncorrected π	Human-Chimpanzee		Human-Macaque		Human-Dog		Human-Mouse	
		Divergence	π	Divergence	π	Divergence	π	Divergence	π
nonPAR	0.000512	0.009814	0.052150	0.049702	0.010297	0.234423	0.002183	0.305070	0.001678
nonPAR$_{minus\_XTR}$	0.000506	0.009782	0.051685	0.049512	0.010212	0.234372	0.002157	0.304460	0.001661
PAR1	0.001505	0.022643	0.066482	0.099892	0.015070	0.337717	0.004457	NA	0.000000
p vs. nonPAR	**0.0000**		*0.0503*		**0.0000**		**0.0000**		*NA*
p vs. nonPAR$_{minusXTR}$	**0.0000**		**0.0368**		**0.0000**		**0.0000**		*NA*
PAR2	0.000643	0.008720	0.073766	0.040967	0.015702	0.218771	0.002940	0.257609	0.002497
p vs. nonPAR	*0.1553*		*0.1416*		*0.1030*		*0.1372*		*0.0993*
p vs. nonPAR$_{minusXTR}$	*0.1429*		*0.1371*		*0.0975*		*0.1298*		*0.0947*
XTR	0.000747	0.010937	0.068256	0.056953	0.013108	0.245717	0.003038	0.336725	0.002217
p vs. nonPAR$_{minusXTR}$	**0.0000**		**0.0000**		**0.0000**		**0.0000**		**0.0000**

Table 1. Diversity across regions of the human X chromosome. Diversity, measured as the average number of pairwise differences per site (π) between the X chromosomes of 26 unrelated genetic females, in each region of the human X chromosome is presented first unnormalized for mutation rate variation, then normalized using human-chimpanzee (hg19-panTro4) divergence and separately normalized for human-macaque (hg19-rheMac3) divergence. The regions analyzed include the pseudoautosomal region 1 (PAR1), pseudoautosomal region 2 (PAR2), X-transposed region (XTR) and the non-pseudoautosomal region either including the XTR

(nonPAR) or excluding the XTR (nonPAR$_{minus\_XTR}$). *P* values from permutation tests with 1000

replicates are shown between each recombining and nonPAR region.

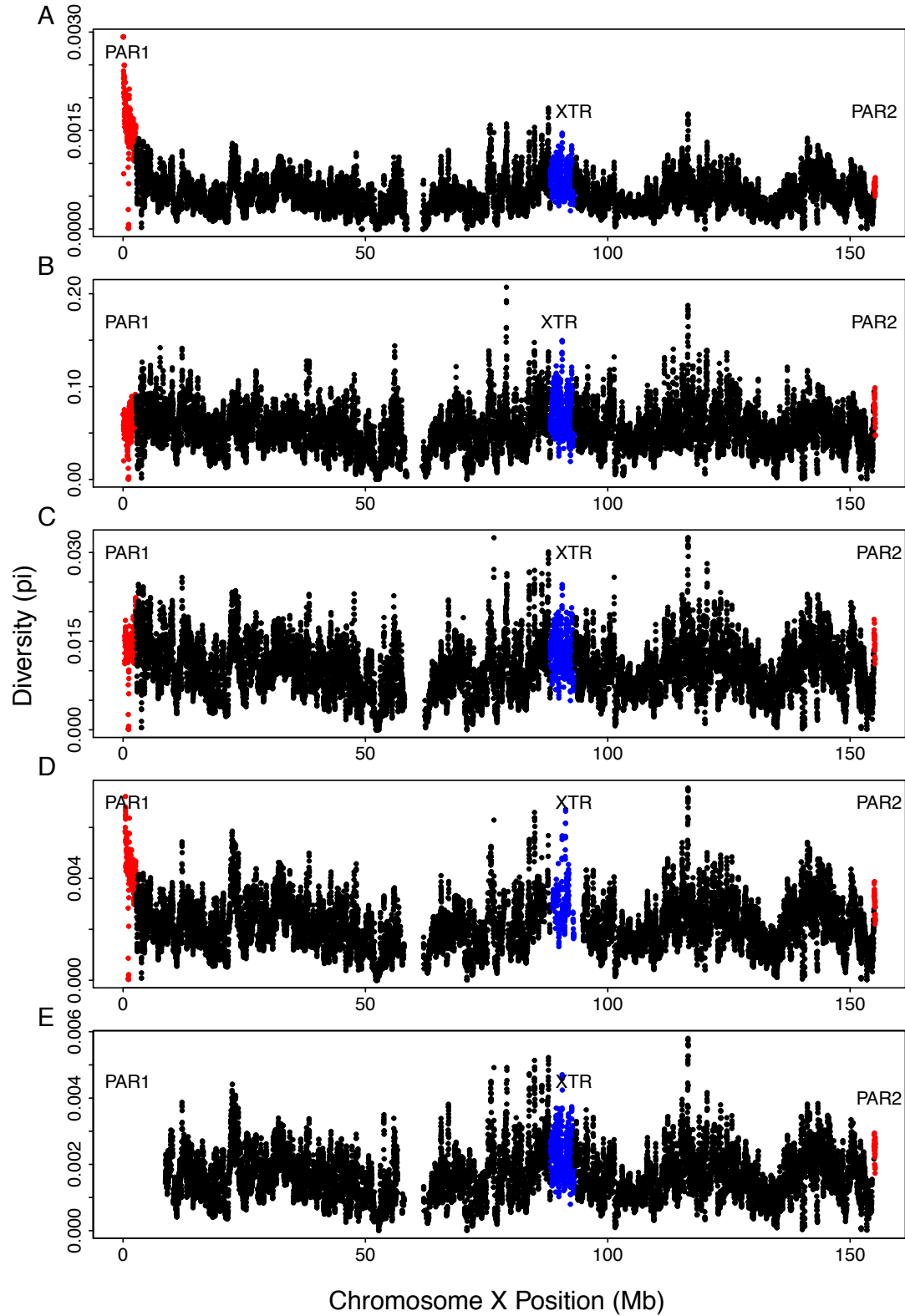## Figures

Figure 1: Diversity along the human X chromosome. Genetic diversity (measured by π) is shown in 100kb overlapping windows across the human X chromosome which includes the pseudoautosomal region 1 (PAR1), shown in red, the non-pseudoautosomal region, shown in black, the X-transposed region (XTR), shown in blue, and the pseudoautosomal region 2 (PAR2), shown in red, for A) human diversity uncorrected for divergence; then human diversity corrected for variable mutation rate using: B) human-chimpanzee divergence; C) human-macaque divergence; D) human-dog divergence; and E) human-mouse divergence.
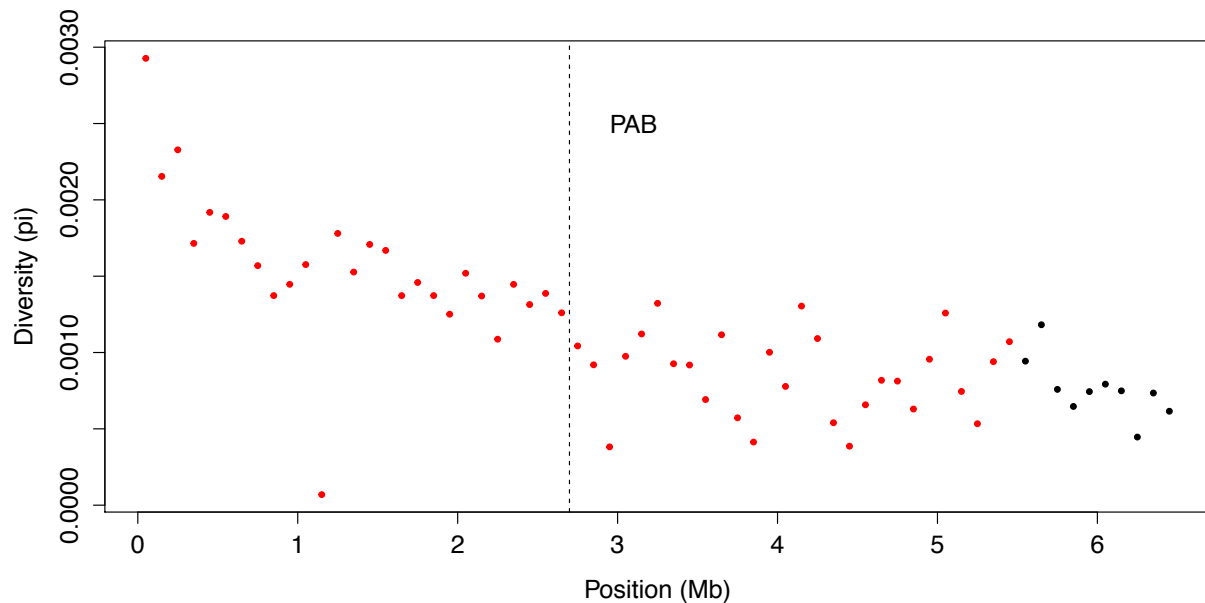
Figure 2: Negative correlation between diversity and distance from Xp, crossing the pseudoautosomal boundary. Diversity in 100kb non-overlapping windows along the psuedoautosomal boundary (PAB) is plotted across the first 6Mb of the human X chromosome, spanning the annotated pseudoautosomal boundary at 2.7Mb. A series of linear regressions were run including 30 windows, sliding by one window across the PAR to the nonPAR region. Each 100kb window is colored red if it is included in a regression in which distance from Xp and diversity are significantly negatively correlated, otherwise the windows are colored black. For the entire region together, diversity is significantly negatively correlated with distance from Xp ($p = 3.281 \times 10^{-10}$; r = -0.7321563) and spans the pseudoautosomal boundary.
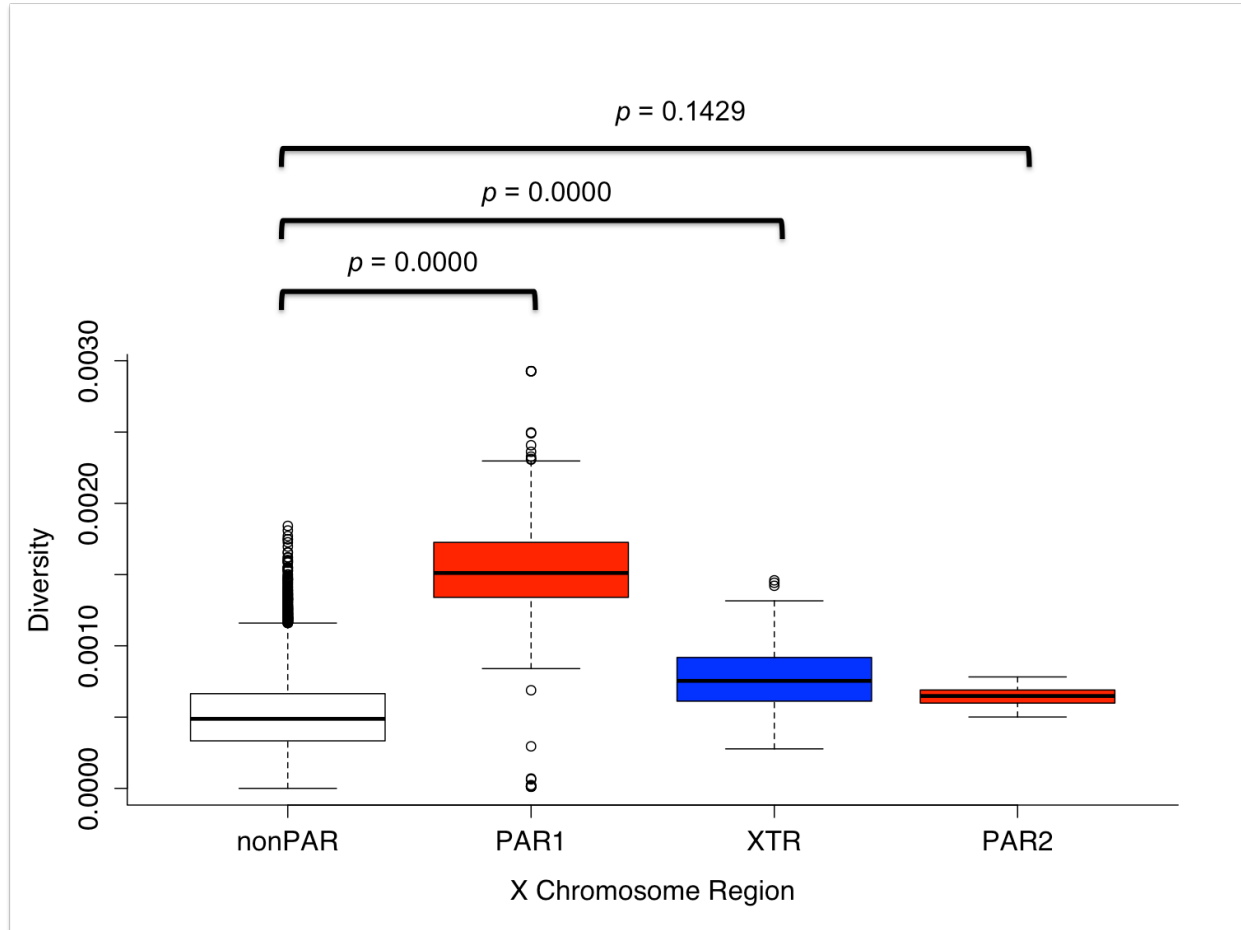
Figure 3: Diversity along the X chromosome split by region. Genetic diversity (measured by π) is shown in boxplots depicting the average diversity with error bars for the nonPAR, PAR1, XTR, and PAR2. The *p value* from a permutation test with 1000 replicates, comparing the diversity of each region to the diversity of the nonPAR are shown.